Stochastic Performance Analysis of a Wireless Finite-State Markov Channel

Kan Zheng, *Senior Member, IEEE,* Fei Liu, Lei Lei, Chuang Lin, *Senior Member, IEEE*, and Yuming Jiang *Member, IEEE*

***Abstract*—Wireless networks are expected to support a diverse range of quality of service requirements and traffic characteristics. This paper undertakes stochastic performance analysis of a wireless finite-state Markov channel (FSMC) by using *stochastic network calculus*. Particularly, delay and backlog upper bounds are derived directly based on the analytical principle behind stochastic network calculus. Both the single user and multi-user cases are considered. For the multi-user case, two channel sharing methods among eligible users are studied, i.e., the *even sharing* and *exclusive use* methods. In the former, the channel service rate is evenly divided among eligible users, whereas in the latter, it is exclusively used by a user randomly selected from the eligible users. When studying the *exclusive use* method, the problem that the state space increases exponentially with the user number is addressed using a novel approach. The essential idea of this approach is to construct a new Markov modulation process from the channel state process. In the new process, the multi-user effect is equivalently manifested by its transition and steady-state probabilities, and the state space size remains unchanged even with the increase of the user number. This significantly reduces the complexity in computing the derived backlog and delay bounds. The presented analysis is validated through comparison between analytical and simulation results.**

***Index Terms*—Finite-state Markov channel, stochastic network calculus, multi-user wireless network, delay bound, backlog bound.**

I. INTRODUCTION

WIRELESS networks are expected to support a diverse range of quality of service requirements and traffic characteristics. However, they have unique characteristics significantly different from those of wired networks, which has made their performance analysis much more challenging. The most fundamental challenge lies in that wireless channels are stochastic in nature. In this paper, we perform stochastic performance analysis of a finite-state Markov channel (FSMC) shared by multiple users.

The finite-state Markov channel (FSMC) model [1] has been widely used for modeling wireless channels. In early 1960s, Gilbert and Elliott proposed a two-state Markov channel with memory, known as the Gilbert-Elliott (GE) channel, to model burst-noise telephone circuits [2] [3]. With the development of wireless networks, this GE model has been adopted for performance studies in wireless fading channel environments. However, modeling a wireless channel as a two-state GE channel has its limitations in cases where channel characteristics may vary dramatically [1] [4]. In [1], the FSMC model was originally proposed to extend the two-state channel model taking into account more channel states. In [4], the FSMC model is shown to be a good model for representing Rayleigh fading channels. The FSMC model has also been used to approximate the block error process of wireless channels [5] [6]. In addition, a wide range of FSMC applications in performance evaluation of wireless networks can be found in the literature, e.g. [7] [8] [9]. Recently, the FSMC model has been used in evaluating cognitive radio, and multiple input and multiple output (MIMO) systems [10] [11]. However, to the best of our knowledge, stochastic service guarantees provided to a user by a FSMC, which are shared by multiple users, have not been well investigated, due largely to the formidable complexity when applying the classical queueing theory [12] [13].

The objective of this paper is to investigate stochastic service guarantees provided by a wireless FSMC. Our primary contribution lies in the derivation of backlog and delay bounds for the system. The approach we use in the derivation follows the analytical principle of a newly-developed theory: stochastic network calculus. It is the stochastic branch of network calculus that has shown to be a good analytical tool for performance analysis of computer networks [14]–[18]. Network calculus can be viewed as a system theory facilitating the derivation of backlog and delay bounds in computer networks. Originally developed for deterministic queueing systems, network calculus has evolved to consider stochastic service guarantees. In recent years, many attempts have been made for the development of stochastic network calculus, including [15], [16], [18]–[30]. Meanwhile, the theory of effective bandwidth, which may be viewed a special case of

Manuscript received February 15, 2012; revised September 13, 2012; accepted October 26, 2012. The associate editor coordinating the review of this paper and approving it for publication was S. Mao.

This work was funded in part by the Program for New Century Excellent Talents in University (NCET-11-0600), the National Basic Research Program of China (973 program: 2010CB328105), and the China NSFC (No.60932003). The work of Yuming Jiang was supported in part by the "Centre for Quantifiable Quality of Service in Communication Systems, Centre of Excellence" appointed by The Research Council of Norway, and funded by The Research Council, Norwegian University of Science and Technology, and UNINETT.

K. Zheng and F. Liu are with the Wireless Signal Processing and Network (WSPN) Lab, Key Laboratory of Universal Wireless Communication, Ministry of Education in Beijing University of Posts & Telecommunications, China (e-mail: kzheng@ieee.org).

L. Lei is with the School of Electrical and Information Engineering, Beijing Jiaotong University, China.

C. Lin is with the Department of Computer Science and Technology, Tsinghua University, China.

Y. Jiang is with the Department of Telematics, Norwegian University of Science and Technology (NTNU), Trondheim, Norway.

Digital Object Identifier 10.1109/TWC.2012.122212.120223

stochastic network calculus [18], has been investigated, which provides a rich variety of probabilistic traffic models based on moment generating functions (MGF) [19]–[21]. The MGF approach has also been used for service modeling and analysis [16], [18], [20], [23], [31]. In addition, the queueing principles underlying stochastic network calculus have been investigated in [18] [26].

In this paper, the analysis is directly based on the analytical principle underlying stochastic network calculus. An upper backlog bound and an upper delay bound are derived for the single-user case. The analysis is then extended to consider the FSMC with multiple users. For the multi-user case, two channel sharing methods among eligible users are investigated. In the even sharing method, the channel service rate is evenly divided among the eligible users, whereas in the exclusive use method, it is exclusively used by a user randomly selected from the eligible user set. Note that, in real wireless networks, the former is not implementable and the latter is more practical. Nevertheless, the former is also analyzed, mainly to demonstrate the exponential increase of the state space in the analysis of the multi-user case. In order to circumvent the resulting formidable computational complexity in calculating the MGF of the service process for the backlog and delay bounds, a novel approach is adopted, of which the essential idea is to construct a new Markov process from the channel state process. In the new process, the multi-user effect is equivalently manifested by the transition and stationary probability, while the state space size remains unchanged even with the increase of the user number. It can significantly reduce the complexity in calculating the obtained backlog and delay bounds. In addition, the constructed new process allows to apply an important result in obtaining the final delay/backlog bounds. To validate the analysis, the analytical bounds on delay and backlog are compared with simulation results, which indicates a close match between them.

The original contributions of this paper are several-fold. First, the considered system is the more general multi-user multi-state Markov channel, of which the single-user or two-user GE channel are special cases. Second, explicit delay and backlog bounds are derived for the system, which do not require heavy traffic assumption or large (infinite) user number assumption. Third, while *even sharing* is studied to elaborate the state explosion problem, the focus is on the more practical *exclusive use* channel sharing method. Particularly, a method dealing with the state explosion problem is adopted for the analysis of *exclusive use* sharing. An important and novel consideration is that the constructed new process allows to apply an important result in obtaining the final delay/backlog bounds. Fourth, physical layer channel parameters are explicitly mapped to the FSMC model and subsequently brought to the obtained delay and backlog bound results. Furthermore, simulation results validating the analysis are presented.

The main contributions of this paper are summarized as follows:

1) The related stochastic network calculus analytical principle is reviewed, which helps make the paper self-contained, facilitating the presentation and easing the reading.
2) Based on the analytical principle, stochastic service analysis of a wireless FSMC, which is shared by multiple users, is conducted. Specifically, analytical delay and backlog bounds for the FSMC are derived.
3) A method to avoid the problem of the exponentially enlarged state space in the analysis is presented.
4) Analytical delay and backlog bounds for the multi-user FSMC are derived and validated by simulations.

The remainder of this paper is organized as follows. Section II reviews related work. In Section III, the system model is introduced. In Section IV, the analytical principle of stochastic network calculus is first introduced, based on which delay and backlog bounds are derived for the single user case. In Section V, delay and backlog bounds are derived for the multi-user case, where both *even sharing* and *exclusive use* sharing are investigated. In Section VI, numerical and simulation results are presented, compared and discussed. Finally, we give further discussion in Section VII and conclude the paper in Section VIII.

II. Related Work

This paper focuses on data traffic performance analysis of a wireless channel, which is shared by multiple users with some scheduling policy. More specifically, we are interested in *probabilistic delay and backlog guarantees* of data traffic under general traffic conditions in such a system. While there is a vast amount of prior literature on the performance of state-aware channel scheduling, most work only addresses heavy traffic behavior or large-buffer asymptotics (e.g., [32]), or in cases of random access scheduling, assumes that the channel behavior seen by a user can be decoupled from other users under a large-user limit (e.g., [33]). Surprisingly, to the best of our knowledge, with respect to data traffic performance analysis of wireless channel scheduling under limited user number and more general traffic conditions, little work has been done to date (see discussion in [34]). One possible reason for this is the lack of a theoretical tool allowing easy and tractable delay/backlog analysis of such systems.

Until most recently, there are few attempts on the study of wireless networks with (stochastic) network calculus. In [35], deterministic network calculus concepts are applied to study the data delay and backlog performance of a Markov channel. However, the system in question has only one user [35]. In the context of stochastic network calculus, the wireless link model is first presented with an on-off impairment model in [36] (see a discussion on this in [37]). However, the focus of [36] is on the general impairment model itself and no specific channel nor scheduling is analyzed in detail.

For the multi-user case, a method based on stochastic network calculus with MGFs is proposed to analyze a two-state Markov channel shared by two users with *even sharing* [12]. Specifically, traffic delay and backlog bounds are presented, which are obtained using the MGF-based stochastic network calculus analysis [23], providing that the state transition probability matrix of the Markov channel is known. However, there is no elaboration on how to find these transition probabilities or how to explicitly relate them to physical layer parameters of the wireless channel. In a recent work [34], which our work is independent from, a queue-length aware scheduling rule is proposed for a multi-channel wireless network where there may be multiple users. However, the focus of [34] is on designing schedule algorithms for multi-channel wireless systems to achieve small-queue performance.

Recently, stochastic network calculus or similar concepts like effective bandwidth and effective capacity have also been applied to performance analysis of other types of (multi-user) wireless networks or systems including MIMO [13] [38] and cognitive radio [39]. However, the considered systems, the channel models and / or the assumptions required for the analysis are different.

Generally, most of the related literature results focus only on the two-state Gilbert-Elliott channel with single user or large number of users where the large user limit allowing to decouple the analysis is applicable. No performance bound of the more general multi-user case for multi-state Markov channel has been well investigated. This may be because the state space of the studied system grows exponentially with the user number, which renders the complexity of the problem intractable. While [12] is perhaps the most related work, the briefly investigated system in [12] is a two-user two-state GE channel case. In addition, in [12], a direct link of physical layer channel parameters to the Markov channel model is missing. Most differently, only *even sharing* is considered in [12] for channel sharing, which, however, is not implementable in real networks. On the other hand, we in this paper consider the more general multi-user FSMC case, exploit an explicit link of physical layer channel parameters to the Markov channel model in the analysis, and investigate both the conceptual *even sharing* and the practical *exclusive use* channel sharing methods.

Related to the stochastic network calculus analytical principle particularly the general concept of service process, *effective capacity*, originally proposed in [31], provides another way to characterize the service provided by a wireless channel. In addition, the general stochastic service curve model, also called statistical service envelope, in stochastic network calculus can also be used to model the service process of a multi-access node (see e.g. [15] [40]), based on which, delay and backlog bounds may further be derived. Indeed, effective capacity can be viewed as a special service model for stochastic network calculus and statistical service curve [15] is indeed an important service model of stochastic network calculus [18]. However, in this paper, we choose to provide analysis directly from the analytical principle and make use of MGF-based stochastic network calculus [23]. Finally, we remark that other analytical methods can be applied to perform delay/backlog performance analysis of wireless systems (see e.g. [41]).

III. System Model

This section introduces the system model, notations and performance metrics. Throughout this paper, matrices and vectors are set in boldface. $\mathbf{1}_N$ denotes $N\times N$ identity matrix. $\mathbb{R}^{M\times N}$ represents the set of $M\times N$ real-value matrix.

A. Finite-State Markov Channel (FSMC)

We consider a wireless FSMC, where the base station transmits information through the channel to mobile stations that are also referred to as users in this paper. The wireless channel is a Rayleigh fading channel. The system is assumed to be work-conserving and within the flow of each user, first-in first-out (FIFO) service is used. The discrete time model with $t \in \mathcal{N}_0 = \{0, 1, \cdots\}$ is adopted.

The channel from the base station to each user is a finite-state Markov channel (FSMC). Under the FSMC model, at any time $t = 1, 2, \ldots$, the channel to a user is described by a set of states $\mathcal{S} = \{s_1, s_2, \cdots, s_L\}$, where L denotes the number of states of the underlying the fading channel [1] [4].

Corresponding to each state $s_l, l \in \{1, 2, \cdots, L\}$ in $\mathcal{S} = \{s_1, s_2, \cdots, s_L\}$, let r_{s_l} denote the service rate of the channel serving workload, which may be achieved by adaptive modulation and coding (AMC) and adjusted according to the wireless channel condition. The rate set, corresponding to L states of the channel, is denoted by $\mathcal{V} = \{r_{s_1}, r_{s_2}, \cdots, r_{s_L}\}$. By convention, the service rate in the worst channel state is set to zero, i.e., $r_{s_1} = 0$. Let $\mathbf{V}(\theta)$ denote the rate diagonal matrix as follows

$$\mathbf{V}(\theta) = \mathrm{diag}(e^{\theta r_1}, e^{\theta r_2}, \cdots, e^{\theta r_L}) \in \mathbb{R}^{L\times L}. \quad (1)$$

Denote by $p_{l,m}$ and π_l the state transition probability from s_l to s_m and the steady-state probability of s_l in $\mathcal{S}$, respectively. Correspondingly, $\mathbf{P} \in \mathbb{R}^{L\times L}$ represents the transition probability matrix with $p_{l,m}, l, m \in \{1, 2, \cdots, L\}$, and $\boldsymbol{\pi} = [\pi_1, \pi_2, \cdots, \pi_L] \in \mathbb{R}^{1\times L}$ is the stationary state distribution vector. The transitions are assumed to occur only between adjacent states, i.e.,

$$p_{l,m} = 0, \forall |l - m| > 1, l, m \in \{1, 2, \cdots, L\}. \quad (2)$$

B. Wireless Channel Parameter Mapping

Let Γ_l, $l \in \{1, \ldots, L+1\}$, be the signal-to-noise ratio (SNR) threshold between the lth and $(l+1)$th states of the Markov model for the user. These threshold values are in increasing order with $\Gamma_{L+1} = \infty$. The channel is in state s_l if the instantaneous SNR γ is between Γ_l and Γ_{l+1}. The stationary probability π_l that the channel is in state s_l is given by

$$\pi_l = \exp(\Gamma_l/\overline{\gamma}) - \exp(\Gamma_{l+1}/\overline{\gamma}), l = 1, \ldots, L. \quad (3)$$

The adjacent-state transition probability over one period of length T_s is calculated as [1] [4]

$$p_{l,l+1} = \frac{\chi(\Gamma_{l+1})T_s}{\pi_l}, \quad l = 1, \ldots, L-1, \quad (4)$$

$$p_{l,l-1} = \frac{\chi(\Gamma_l)T_s}{\pi_l}, \quad l = 2, \ldots, L, \quad (5)$$

where $\chi(\Gamma_l)$ denotes the level cross rate at an instantaneous SNR for the user. Under the assumption of a Rayleigh fading channel, $\chi(\Gamma_l)$ is expressed as [4]

$$\chi(\Gamma_l) = \sqrt{\frac{2\pi\Gamma_l}{\overline{\gamma}}} f_D \exp(-\frac{\Gamma_l}{\overline{\gamma}}),$$

where f_D denotes the mobility-induced Doppler spread of the fading channel, and $\overline{\gamma} = \mathbb{E}[\gamma]$ is the average received SNR.

With (2), (4) and (5), $p_{l,l}$ can be derived from the normalization condition $\sum_{m=1}^{L} p_{l,m} = 1$ as

$$p_{l,l} = \begin{cases} 1 - p_{l,l+1} - p_{l,l-1}, & (l = 2, \ldots, L-1) \\ 1 - p_{l,l+1}, & (l = 1) \\ 1 - p_{l,l-1}, & (l = L) \end{cases} . \quad (6)$$

C. Performance Metrics

For each user, let us define real-valued cumulative functions $A(0,t)$ and $D(0,t)$ as the (data traffic) arrival and departure processes of the channel, respectively. Specifically, the arrival process $A(0,t)$ represents the amount of workload arriving at the mobile station/channel in the interval of $(0,t]$, and the departure process $D(0,t)$ indicates the amount of workload departing from the channel and reaching the corresponding mobile station in the interval of $(0,t]$. Define that $A(s,t) := A(0,t) - A(0,s)$ and $D(s,t) := D(0,t) - D(0,s)$. By definition, $A(0,t)$ and $D(0,t)$ are non-negative and non-decreasing with t, and $A(t,t) = 0$, $D(t,t) = 0, \forall t \geq 0$.

To conduct the performance analysis of the network, we are particularly interested in two metrics, namely backlog and delay. The backlog at time $t \geq 0$ is given by

$$b(t) = A(0,t) - D(0,t) \quad (7)$$

and the delay at time $t \geq 0$ is defined as

$$d(t) = \inf[\tau \geq 0 : A(0,t) \leq D(0,t+\tau)]. \quad (8)$$

IV. FSMC with Single User

In this section, we assume there is only one mobile station or user. We analyze the stochastic performance guarantees provided by the finite state Markov channel to the user. The adopted approach follows the analytical principle underlying stochastic network calculus [18], [23], [26].

A. Stochastic Network Calculus Basics: Analytical Principle and Bounds

This subsection reviews the analytical principle underlying stochastic network calculus, directly based on which we further present how delay and backlog bounds can be derived. We highlight that the content in this subsection can be mostly found from the stochastic network calculus literature [18], [23], [26]. However, to ease the presentation and make the paper self-contained, we choose to skip the traffic modeling and service modeling parts[1], base the analysis directly on the traffic process and the service process, and provide the key analysis steps.

In general, for a single user system, the following inequality holds

$$b(t+1) \leq \max\{b(t) + A(t,t+1) - S(t,t+1), 0\}, \quad (9)$$

where $S(t,t+1)$ denotes a lower bound on the amount of workload that can be served by the system in $(t,t+1]$ to the user. In the rest of this paper, $S(t,t+1)$ satisfying (9) is taken as the service rate of the channel at time $t+1$ to the user. When the equality of (9) is true, the relation is also known as the Lindley recursion that governs work-conserving queueing systems. Through applying (9) iteratively, we arrive at the following relationship for backlog

$$b(t) \leq \max_{0 \leq s \leq t} \{A(s,t) - S(s,t)\} \quad (10)$$

and hence

$$P\{b(t) > b\} \leq P\{\max_{0 \leq s \leq t} \{A(s,t) - S(s,t)\} > b\}. \quad (11)$$

Recall that $b(t) = A(0,t) - D(0,t)$. So we have

$$D(0,t) = A(0,t) - b(t) \geq \inf_{0 \leq s \leq t} \{A(0,s) + S(s,t)\}. \quad (12)$$

For delay, its definition, $d(t) = \inf\{\tau : A(0,t) \leq D(0,t+\tau)\}$, implies that for any $\tau \geq 0$, if $d(t) > \tau$, $A(0,t) > D(0,t+\tau)$ is true, since otherwise if $A(0,t) \leq D(0,t+\tau)$ and $d(t) \leq \tau$, it would contradict the condition $d(t) > \tau$. In other words, event $\{d(t) > \tau\}$ implies event $\{A(0,t) > D(0,t+\tau)\}$, and hence

$$P\{d(t) > \tau\} \leq P\{A(0,t) > D(0,t+\tau)\} \quad (13)$$

with which and (12), the following relation for delay holds

$$P\{d(t) > \tau\} \leq P\{\max_{0 \leq s \leq t} \{A(s,t) - S(s,t+\tau)\} > 0\}. \quad (14)$$

Assume that the arrival process $A(0,t)$ and the service bound process $S(0,t)$ are two statistically independent, stationary random processes throughout this paper. Then, the distribution of $A(s,s+t)$ equals that of $A(0,t)$ for all $s > 0$ due to the stationarity.

For all real θ, the moment generating functions of $A(0,t)$ and $S(0,t)$ are defined as

$$\begin{aligned} \mathrm{M}_A(\theta,t) &= E[e^{\theta A(0,t)}], \\ \mathrm{M}_S(\theta,t) &= E[e^{\theta S(0,t)}], \end{aligned} \quad (15)$$

where $E[x]$ denotes the expectation of random process x. For notation purposes, we also adopt

$$\overline{\mathrm{M}}_S(\theta,t) = \mathrm{M}_S(-\theta,t).$$

Applying the Chernoff bound to the right hand sides of both (11) and (14), we obtain for backlog that

$$\begin{aligned} P & \quad \{\max_{0 \leq s \leq t} \{A(s,t) - S(s,t)\} > b\} \\ \leq & \quad e^{-\theta b} E[e^{\theta \max_{0 \leq s \leq t}\{A(s,t) - S(s,t)\}}] \\ \leq & \quad e^{-\theta b} \sum_{s=0}^{t} E[e^{\theta A(s,t) - \theta S(s,t)}] \\ = & \quad e^{-\theta b} \sum_{s=0}^{t} E[e^{\theta A(0,s) - \theta S(0,s)}] \\ \leq & \quad e^{-\theta b} \sum_{s=0}^{\infty} \mathrm{M}_A(\theta,s) \overline{\mathrm{M}}_S(\theta,s) \end{aligned} \quad (16)$$

[1]They are often introduced in other papers when stochastic network calculus is applied, but explaining these models can be lengthy and understanding them highly demanding.

and similarly for delay that

$$\begin{aligned}
&P\{\max_{0\le s\le t}\{A(s,t)-S(s,t+\tau)\}>0\}\\
&\le E[e^{\theta\max_{0\le s\le t}\{A(s,t)-S(s,t+\tau)\}}]\\
&\le \sum_{s=0}^{t} E[e^{\theta A(s,t)-\theta S(s,t+\tau)}]\\
&\le \sum_{s=\tau}^{\infty} \mathrm{M}_A(\theta,s-\tau)\overline{\mathrm{M}}_S(\theta,s) \qquad (17)
\end{aligned}$$

for $\forall\theta>0$.

The following lemma summarizes the above discussions.

Lemma 1: Assume $S(0,t)$ and $A(0,t)$ are statistically independent and stationary with MGFs $\mathrm{M}_S(\theta,t)$ and $\mathrm{M}_A(\theta,t)$, respectively. Then, the backlog and delay are bounded by[2]

$$P\{b(t)>b\} \le e^{-\theta b}\sum_{s=0}^{\infty}\mathrm{M}_A(\theta,s)\overline{\mathrm{M}}_S(\theta,s) \qquad (18)$$

$$P\{d(t)>d\} \le \sum_{s=d}^{\infty}\mathrm{M}_A(\theta,s-d)\overline{\mathrm{M}}_S(\theta,s) \qquad (19)$$

for $\forall\theta>0$.

Lemma 1 suggests that if the MGFs of the arrival process $A(0,t)$ and the lower bound service process $S(0,t)$ are known, the backlog and delay bounds are readily obtained by using the lemma. In this paper, we assume the MGF of $A(0,t)$ is given, and our focus is on the MGF of $S(0,t)$. For this, we will heavily rely on the following result, which has been proved in the literature (e.g., see [21]).

Lemma 2: Suppose $\mathcal{S}(t)$ is a homogeneous Markov process of states $\{s_1,s_2,\cdots,s_L\}$ with transition matrix $\mathbf{P}$ and stationary state distribution vector $\boldsymbol{\pi}$. Then, for the Markov modulated process $r(t)=r_{\mathcal{S}(t)}(t)$, the MGF of its cumulative process $S(0,t)=\sum_{\tau=1}^{t}r(\tau)$ is given by

$$\mathrm{M}_S(\theta,t)=\boldsymbol{\pi}(\mathbf{V}(\theta)\mathbf{P})^{t-1}\mathbf{V}(\theta)\mathbf{1}_L. \qquad (20)$$

where $\mathbf{V}(\theta)$ is a diagonal matrix with $\mathbf{V}(\theta)=\mathrm{diag}(\mathrm{M}_{r_{s_1}}(\theta),\mathrm{M}_{r_{s_2}}(\theta),\ldots,\mathrm{M}_{r_{s_L}}(\theta))$ for $\forall\theta$, and $\mathbf{1}_L$ denotes a column vector with all its L elements being one.

B. Backlog and Delay Bounds for Single-User FSMC

Let $\mathcal{S}(t)$ be the channel state process, where $\mathcal{S}(t)=s_l$ if the channel is in state s_l at time t. Recall that the channel is a FSMC characterized by the state transition probability matrix $\mathbf{P}\in\mathbb{R}^{L\times L}$, whose elements $p_{l,m}$ $\forall l,m\in\{1,2,\ldots,L\}$ are decided from (2) – (6), and the stationary state distribution vector $\boldsymbol{\pi}=[\pi_1,\pi_2,\cdots,\pi_L]\in\mathbb{R}^{1\times L}$, whose elements π_l are found by (3). The channel state process $\mathcal{S}(t)$ is hence a homogeneous Markov process with state transition probability matrix $\mathbf{P}$ and stationary state distribution vector $\boldsymbol{\pi}$.

Note that the service rate in state s_l is $r_{s_l}, l\in\{1,2,\cdots,L\}$. This implies that the service rate process $r_{\mathcal{S}(t)}(t)$ is a Markov modulated process and Lemma 2 is hence applicable. Let $S(t-1,t)=r_{\mathcal{S}(t)}(t)$, $t=1,2,\ldots$, and then $S(0,t)=\sum_{\tau=1}^{t}S(\tau-1,\tau)=\sum_{\tau=1}^{t}r_{\mathcal{S}(\tau)}(\tau)$. From Lemma 2, we can then obtain $\mathrm{M}_S(\theta,t)$.

Suppose the allowable backlog violation probability is $\epsilon\in(0,1]$. Then, letting the right hand side of (18) equal ϵ and with some mathematical manipulation, a backlog bound can be obtained. Similarly, a delay bound can also be found. The following theorem presents the obtained bounds, where (21) and (22) are consistent with stochastic network calculus results in the literature, e.g., [23].

Theorem 1: For the FSMC with a single user, if the arrival process $A(0,t)$ of the user is stationary and independent of the service process $S(0,t)$ provided by the channel, then an upper backlog bound and an upper delay bound, each with at most violation probability $\epsilon\in(0,1]$, are given by[3]

$$b=\inf_{\theta>0}[\frac{1}{\theta}(\ln\sum_{s=0}^{\infty}\mathrm{M}_A(\theta,s)\overline{\mathrm{M}}_S(\theta,s)-\ln\varepsilon)], \qquad (21)$$

$$d=\inf_{\theta>0}\{\inf[\tau:\frac{1}{\theta}(\ln\sum_{s=\tau}^{\infty}\mathrm{M}_A(\theta,s-\tau)\overline{\mathrm{M}}_S(\theta,s)-\ln\varepsilon)\le 0]\} \qquad (22)$$

with $\overline{\mathrm{M}}_S(\theta,0)=1$ and for $s\ge 1$,

$$\overline{\mathrm{M}}_S(\theta,s)=\boldsymbol{\pi}(\mathbf{V}(-\theta)\mathbf{P})^{s-1}\mathbf{V}(-\theta)\mathbf{1}_L, \qquad (23)$$

where $\mathbf{P}$ is determined by (2) – (6), $\boldsymbol{\pi}$ by (3) and $\mathbf{V}(-\theta)=\mathrm{diag}(e^{-\theta r_1},\cdots,e^{-\theta r_L})$.

V. FSMC WITH MULTIPLE USERS

We now extend the analysis to consider the wireless channel shared by N multiple users. It is assumed that these N users experience independent and statistically identical fading channels. For each user, the channel is modeled as a FSMC with L states as in the single-user case described in the previous section.

Specifically, the state space of the multi-user channel becomes $\tilde{\mathcal{S}}=\{\tilde{s}_1,\cdots,\tilde{s}_{L^N}\}$ with L^N states. The kth state in this multi-user channel consists of N users' states, i.e.,

$$\tilde{s}_k:=\langle s_{u_{k,1}},s_{u_{k,2}},\cdots,s_{u_{k,n}},\cdots,s_{u_{k,N}}\rangle, 1\le k\le L^N,$$

where $s_{u_{k,n}}\in\mathcal{S}$ has its corresponding rate $r_{s_{u_{k,n}}}\in\mathcal{V}$, and $u_{k,n}, 1\le u_{k,n}\le L$, is the state index of user n, $1\le n\le N$, when the multi-user channel is in state $\tilde{s}_k\in\tilde{\mathcal{S}}$. For all users, they have the same $\mathcal{S}$ and $\mathcal{V}$ which are also the same as in the single user case, defined in the previous section.

The state transition probability from $\tilde{s}_k$ to $\tilde{s}_j$ and the steady-state probability of $\tilde{s}_k$ in the multi-user channel can be computed by

$$\tilde{p}_{\tilde{s}_k,\tilde{s}_j}=\prod_{n=1}^{N}p_{s_{u_{k,n}},s_{u_{j,n}}}, k,j\in\{1,2,\cdots,L^N\}, \qquad (24)$$

and

$$\tilde{\pi}_{\tilde{s}_k}=\prod_{n=1}^{N}\pi_{s_{u_{k,n}}}, k\in\{1,2,\cdots,L^N\}, \qquad (25)$$

respectively, where for each user n, the transition probability $p_{s_{u_{k,n}},s_{u_{j,n}}}$ is the same as in the single user case and determined by (2) – (6), and the steady-state probability $\pi_{s_{u_{k,n}}}$ is

[2]While not explicitly shown, inequalities (18) and (19) can also be derived from results in [23].

[3]We remark that (21) and (22) are also seen in [23], where discussion on their convergence / stability can be found.

also the same as in the single user case and determined by (3).

Let us denote the steady-state probability vector and the transition probability matrix in the multi-user channel by

$$\tilde{\boldsymbol{\pi}} = [\tilde{\pi}_1, \tilde{\pi}_2, \cdots, \tilde{\pi}_{L^N}] \in \mathbb{R}^{1 \times L^N}$$

and

$$\tilde{\mathbf{P}} \in \mathbb{R}^{L^N \times L^N}$$

respectively, with $\tilde{p}_{\tilde{s}_k, \tilde{s}_j}, k, j \in \{1, 2, \cdots, L^N\}$.

Denote by $\tilde{\mathcal{S}}(t)$ the process of the N users' channel states. We then have $\tilde{\mathcal{S}}(t) = \tilde{s}_k$, if the system is in state $\tilde{s}_k$ at time t. Since the N users experience independent and statistically identical channels and for each user the channel is a FSMC, it can be verified that $\tilde{\mathcal{S}}(t)$ is a homogeneous Markov process, which means the transition probability matrix is time-invariant according to (24). In particular, $\tilde{\mathcal{S}}(t)$ has the state transition probability matrix $\tilde{\mathbf{P}}$ and stationary state distribution vector $\tilde{\boldsymbol{\pi}}$.

The performance of the multi-user FSMC depends on the scheduling strategy in use. Channel-aware algorithms aim at improving the scheduling performance by utilizing the channel state information. In this paper, the greedy algorithm [42], referred to as the Max-SNR (maximum signal to noise ratio) or Maximum Rate Scheduling (MRS) algorithm, is employed. This algorithm always selects the user(s) with the best channel condition for transmission at a given time instant. Scheduling users according to their channel state can explore the multi-user diversity and provide significant performance gains due to the independence of fading statistics across users. It has been shown that this greedy algorithm is optimal in the sense of capacity-approaching [42].

When there are several users having the same best channel conditions, two channel sharing methods, namely *even sharing* and *exclusive use*, are considered, which will be analyzed separately in the next subsections.

We highlight that *even sharing* is a conceptual channel sharing method and not implementable in real networks. One main reason of including *even sharing* in this paper is that it is highly intuitive. Another reason is that it helps us understand the state space exploration challenge in computing the obtained bounds, which motivates us to search for a way to address a similar challenge in analyzing the more practical *exclusive use* channel sharing method.

A. Method 1: Even Sharing

In this sharing method, we assume that the service rate corresponding to their best channel state is evenly divided among the eligible users that are in the best channel state as in [12]. For the sake of analysis, the channels allocated to the eligible users are assumed to be orthogonal, implying that no interference exists among users. Then, the assumption of independence among users holds for the even sharing method. A *straightforward* approach for the analysis is to extend the single-user case analysis detailed in the previous section.

At time t, the maximum rate provided by the channel to any user is denoted by $\max(r_{s_{u_{k,1}}}(t),\ r_{s_{u_{k,2}}}(t), \cdots, r_{s_{u_{k,N}}}(t))$. Let

$$\mathcal{M}(t) = \{m|\ r_{s_{u_{k,m}}}(t) = \max(r_{s_{u_{k,1}}}(t), r_{s_{u_{k,2}}}(t), \cdots, r_{s_{u_{k,N}}}(t))\}$$

be the set of eligible users with the best channel conditions providing the maximum rate, and $||\mathcal{M}(t)||$ be the size of set $\mathcal{M}(t)$.

With the even sharing method, the service rate to user n, which can be provided in state $\tilde{s}_k \in \tilde{\mathcal{S}}$ at time t, is given by

$$\tilde{r}_{n,\tilde{s}_k}(t) = \begin{cases} \frac{1}{||\mathcal{M}(t)||} r_{s_{u_{k,n}}}(t) & \text{if } n \in \mathcal{M}(t) \\ 0, & \text{otherwise} \end{cases}, \tag{26}$$

where the maximum rate is evenly divided among the eligible users.

For any user n, let $S(t-1,t) = \tilde{r}_{n,\tilde{\mathcal{S}}(t)}(t)$, $t = 1, 2, \ldots$, and hence $S(0,t) = \sum_{\tau=1}^{t} S(\tau-1,\tau) = \sum_{\tau=1}^{t} \tilde{r}_{n,\tilde{\mathcal{S}}(\tau)}(\tau)$. Then, $\tilde{r}_{n,\tilde{\mathcal{S}}(t)}(t)$, $t = 1, 2, \ldots$, also forms a Markov modulated process, and the MGF of $S(0,t)$ can be readily found by Lemma 2. Particularly for any user n, the MGF of its process $S(0,t)$ is

$$\mathrm{M}_S(\theta, t) = \tilde{\boldsymbol{\pi}}(\tilde{\mathbf{V}}_n(\theta)\tilde{\mathbf{P}})^{t-1}\tilde{\mathbf{V}}_n(\theta)\mathbf{1}_{L^N}. \tag{27}$$

with the corresponding rate matrix

$$\tilde{\mathbf{V}}_n(\theta) = \mathrm{diag}(e^{\theta\tilde{r}_{n,\tilde{s}_1}}, e^{\theta\tilde{r}_{n,\tilde{s}_2}}, \cdots, e^{\theta\tilde{r}_{n,\tilde{s}_{L^N}}}) \in \mathbb{R}^{L^N \times L^N}.$$

Since all the users are assumed to have independent and statistically identical channels, the MGFs of the service processes to different users are thus equal. As a result, the user index n can be dropped from $\mathrm{M}_S(\theta, t)$ as well as for $\tilde{\mathbf{V}}(\theta)$.

Similar to the single-user analysis, a delay bound and a backlog bound for the multi-user FSMC with the even sharing method can be obtained from Lemmas 1 and 2, which are consistent with the results presented in [12] for a two-user two-state GE channel:

Theorem 2: Consider the FSMC with multiple users. If for any user, the arrival process $A(0,t)$ of the user is stationary and independent of service process $S(0,t)$ that can be provided by the channel to the user, then an upper backlog bound and an upper delay bound, each with at most violation probability $\epsilon \in (0,1]$, are given by

$$b = \inf_{\theta>0}[\frac{1}{\theta}(\ln\sum_{s=0}^{\infty}\mathrm{M}_A(\theta,s)\overline{\mathrm{M}}_S(\theta,s) - \ln\varepsilon)],$$

$$d = \inf_{\theta>0}\{\inf[\tau : \frac{1}{\theta}(\ln\sum_{s=\tau}^{\infty}\mathrm{M}_A(\theta,s-\tau)\overline{\mathrm{M}}_S(\theta,s) - \ln\varepsilon) \leq 0]\}$$

with $\overline{\mathrm{M}}_S(\theta, 0) = 1$ and for $s \geq 1$

$$\overline{\mathrm{M}}_S(\theta, s) = \tilde{\boldsymbol{\pi}}(\tilde{\mathbf{V}}(-\theta)\tilde{\mathbf{P}})^{s-1}\tilde{\mathbf{V}}(-\theta)\mathbf{1}_{L^N}, \tag{28}$$

where $\tilde{\mathbf{P}}$ and $\tilde{\boldsymbol{\pi}}$ are given by (24) and (25), respectively, and $\tilde{\mathbf{V}}(-\theta) = \mathrm{diag}(e^{-\theta\tilde{r}_{n,\tilde{s}_1}}, e^{-\theta\tilde{r}_{n,\tilde{s}_2}}, \cdots, e^{-\theta\tilde{r}_{n,\tilde{s}_{L^N}}})$.

However, the above straightforward approach for analyzing the performances of the even-sharing method leads to an exponentially increased state space, i.e., from L states for the single user case to L^N-states for the multi-user case. Particularly, for the single user case as shown in (23), the MGF of the service process is determined by its stationary state distribution vector $\boldsymbol{\pi}$, its transition probability matrix $\mathbf{P}$, and its service rate matrix $\mathbf{V}(\theta)$. For the multi-user case discussed in this subsection, the MGF of the service process of each user is determined by (28). Although the service rate under each of the multi-user channel states is within the rate set $\mathcal{V}$ with L possible rates, there are altogether L^N possible channel states. This results in extraordinary high computational complexity in calculating the MGF of the service process.

B. Method 2: Exclusive Use

In the even sharing method, we have assumed that the maximum service rate that the channel can provide at a time is evenly divided among all the eligible users that have the same best channel condition at the given time. However, in real wireless networks, it is typical that only one user uses the channel exclusively at any time. In this subsection, we analyze the multi-user case, where a greedy maximum rate scheduling algorithm is used in conjunction with exclusive channel use. More specifically, when there are multiple eligible users, one of them is randomly selected to use the channel exclusively.

As discussed in the previous subsection, the state space of multi-user cases will be increased exponentially if only to extend the single-user case analysis to multi-user cases. To tackle this challenge, we adopt a new approach to analyze the multi-user case with exclusive channel use. The essential idea is to construct a new process with a reduced state space from $\tilde{\mathcal{S}} = \{\tilde{s}_1, \cdots, \tilde{s}_{L^N}\}$ with L^N states. This new process can reflect the exclusive use effect in the multi-user case, while keeping the rate matrix unchanged as in the single-user case. The construction of the new process will be described in the following.

Since the channel processes of all the N users are independent and identically distributed, each user has the same statistical characteristics. Without loss of generality, user n is taken for the discussions below, while the analysis is same for all the other users.

First, we construct a new homogeneous process $\mathcal{X}(t)$, $t = 1, 2, \ldots$ with state space $\mathcal{X} = \{x_1, x_2, \cdots, x_L\}$ for user n, which considers the effects of both channel varying characteristics and the employed multi-user scheduling algorithm. In this newly constructed process, at any time t, user n will be in state $x_l \in \mathcal{X}$, i.e., $\mathcal{X}(t) = x_l$, only if the user is in channel state $s_l \in \mathcal{S}, 1 \leq l \leq L$, and it is also selected to use the channel at that time. The corresponding channel service rate of user n in state x_l is $r_{x_l} = r_{s_l} \in \mathcal{V}$. However, when a user is not selected by the scheduler, it will always be in state x_1 with the corresponding channel service rate of $r_{x_1} = 0$, irrespective of the user's channel state. Due to this, the new service process has the same corresponding rate set $\mathcal{V}$ and rate matrix $\mathbf{V}(\theta)$ as those in the single-user case.

Next, let us discuss the stationary state probability vector

$$\breve{\boldsymbol{\pi}} = [\breve{\pi}_{x_1}, \breve{\pi}_{x_2}, \cdots, \breve{\pi}_{x_L}] \in \mathbb{R}^{1\times L}. \tag{29}$$

The steady-state probability of user n in state x_l can be computed from $\breve{\boldsymbol{\pi}}$ according to Bayes' theorem as

$$\breve{\pi}_{x_l} = \sum_{k=1}^{L^N} \tilde{\pi}_{\tilde{s}_k} P_n(x_l|\tilde{s}_k), 1 \leq l \leq L \tag{30}$$

where $\tilde{\pi}_{\tilde{s}_k}$ is the steady-state probability of state $\tilde{s}_k$ in the multi-user channel as given in (25), and $P_n(x_l|\tilde{s}_k)$ represents the probability that user n is in state x_l, providing that the multi-user channel state is $\tilde{s}_k$.

As stated above, if a user is in the worst channel state s_1 or not being scheduled, the user will not be assigned any resource for transmission and it will be in state x_1 of the newly constructed process. Therefore, there are three distinct cases where user n is in state x_1, when the multi-user channel state is $\tilde{s}_k$:

- *Case 1*: When user n is not one of the eligible users, i.e., $n \notin \mathcal{M}$, it is in state x_1;
- *Case 2*: If all the users are in the worst channel state s_1, user n will be in state x_1 even if it is scheduled, i.e., $n \in \mathcal{M}$; and
- *Case 3*: If user n is in the currently best channel state and hence one of the eligible users, i.e., $n \in \mathcal{M}$, but it is not scheduled, it will be in state x_1.

With the exclusive-use Max-SNR scheduling algorithm, the probability that user n is in state x_1 given $\tilde{s}_k, 1 \leq k \leq L^N$, is computed by

$$P_n(x_1|\tilde{s}_k) = \begin{cases} 1, & \text{if } \ n \notin \mathcal{M} \\ 1, & \text{if } \ n \in \mathcal{M}, \ r_{s_{k,u_n}} = r_1 = 0 \\ 1 - \frac{1}{||\mathcal{M}||} & \text{if } \ n \in \mathcal{M}, \ r_{s_{k,u_n}} > r_1 = 0 \\ 0, & \text{otherwise.} \end{cases} \tag{31}$$

For any other state x_l, the probability that user n is in that state given $\tilde{s}_k$ can be computed by

$$P_n(x_l|\tilde{s}_k) = \begin{cases} \frac{1}{||\mathcal{M}||} & \text{if } \ n \in \mathcal{M}, \ r_{s_{k,u_n}} = r_{x_l} > 0 \\ 0, & \text{otherwise} \end{cases}, \tag{32}$$

$$1 < l \leq N, 1 \leq k \leq L^N$$

At this point, we would like to highlight that $\tilde{\mathcal{S}}(t)$ with state space $\tilde{\mathcal{S}}$ is a multi-dimensional Markov process, i.e., $\tilde{\mathcal{S}}(t) = [\tilde{\mathcal{S}}_1(t), \ldots, \tilde{\mathcal{S}}_N(t)]$, where $\tilde{\mathcal{S}}_n(t)$ denotes the state process of user n. However, the constructed process $\mathcal{X}(t)$ with state space $\mathcal{X} = \{x_1, x_2, \cdots, x_L\}$, which is essentially the same as the state space $\{s_1, s_2, \cdots, s_L\}$, is not Markovian. This prevents one from applying Lemma 2 to our later analysis, which is crucial.

In the following, a technique to circumvent the state explosion problem [43] is adopted. Particularly, in [34], a similar technique is used to study processes closely related to the $\mathcal{X}(t)$ type. More specifically, define process $\mathcal{Y}(t)$ as follow

$$\mathcal{Y}(t) = \max_{1 \leq n \leq N} \tilde{\mathcal{S}}_n(t). \tag{33}$$

Note that $\mathcal{Y}(t)$ also has the state space of $\{s_1, s_2, \cdots, s_L\}$, and $\mathcal{X}(t)$ is produced from $\mathcal{Y}(t)$.

While $\mathcal{Y}(t)$ is not Markovian either, the existing result (see Theorem 2 in [34]) shows that process $\mathcal{Y}(t)$ has a stationary distribution that can be given by a homogeneous Markov process. Denote by $\mathcal{Y}'(t)$ a homogeneous Markov process with state space $\{s_1, s_2, \cdots, s_L\}$, whose transition probabilities are given by

$$P\{\mathcal{Y}'(t+1) = s_m|\mathcal{Y}'(t) = s_l\} = P\{\mathcal{S}(t+1) \in \tilde{s}_m|\mathcal{S}(t) \in \tilde{s}_l\}$$

where

$$\tilde{s}_l := \left\{ \langle s_{u_1}, \ldots, s_{u_N} \rangle : \max_{1 \leq n \leq N} s_{u_n} = s_l \right\}.$$

Then, $\mathcal{Y}(t)$ and $\mathcal{Y}'(t)$ follow the same stationary distribution [34]. This result suggests a way to calculate the stationary distribution of $\mathcal{Y}(t)$. That is, we can treat $\mathcal{Y}(t)$ as if it were Markovian. This method, which has been adopted in [34], is capable of substantially reducing the computational complexity.

Following the same technique, we construct a Markov process $\mathcal{X}'(t)$ corresponding to $\mathcal{X}(t)$, which has the same state space $\{x_1, x_2, \cdots, x_L\}$ as $\mathcal{X}(t)$, and its transition probability matrix is

$$\breve{\mathbf{P}} \in \mathbb{R}^{L\times L}$$

where $\breve{p}_{u,v}, u, v \in \{1, 2, \cdots, L\}$, denotes the state transition probability from state x_u to state x_v. It is calculated by

$$\breve{p}_{u,v} = \sum_{k=1}^{L^N} \{P_n(\tilde{s}_k|x_u) \sum_{l=1}^{L^N} \tilde{p}_{\tilde{s}_k,\tilde{s}_l} P_n(x_v|\tilde{s}_l)\}, 1 \le u, v \le L \tag{34}$$

where $P_n(\tilde{s}_k|x_u)$, the probability of user n in the multi-user state $\tilde{s}_k$ given state x_u, can be expressed as

$$P_n(\tilde{s}_k|x_u) = \frac{\tilde{\pi}_{\tilde{s}_k} P_n(x_u|\tilde{s}_k)}{\breve{\pi}_{x_u}}. \tag{35}$$

Fortunately, it is easily found that the Markov process $\mathcal{X}'(t)$ has the same steady-state distributions as $\mathcal{X}(t)$. Since the statistical property of a random process is largely determined by its steady-state distribution, we can treat $\mathcal{X}(t)$ as if it were a Markov process in the analysis. Alternatively, we can approximate $\mathcal{X}(t)$ with $\mathcal{X}'(t)$, and apply $\mathcal{X}'(t)$ to the analysis.

Specifically, for user n, we approximate its rate process $r_{n,\mathcal{X}(t)}(t)$ as $r_{n,\mathcal{X}'(t)}(t)$. Let $S(t-1,t) = r_{n,\mathcal{X}'(t)}(t)$, $t = 1, 2, \ldots$, and hence $S(0,t) = \sum_{\tau=1}^{t} S(\tau-1,\tau) = \sum_{\tau=1}^{t} r_{n,\mathcal{X}'(\tau)}(\tau)$. Then, $r_{n,\mathcal{X}'(t)}(t)$,$t = 1, 2, \ldots$, forms a Markov modulated process. The MGF of process $S(0,t)$ is now readily attainable by Lemma 2. For user n, the MGF of $S(0,t)$ becomes (approximately)

$$\mathrm{M}_S(\theta,t) = \breve{\boldsymbol{\pi}}(\mathbf{V}(\theta)\breve{\mathbf{P}})^{t-1}\mathbf{V}(\theta)\mathbf{1}_L. \tag{36}$$

Then, similar to the single-user analysis, a delay bound and a backlog bound for the exclusive-use multi-user FSMC can be finally obtained through the following theorem.

Theorem 3: Consider the FSMC with multiple users. If for any user, the arrival process $A(0,t)$ of the user is stationary and independent of the service process $S(0,t)$ that can be provided by the channel to the user, then an upper backlog bound and an upper delay bound, each with at most violation probability $\epsilon \in (0,1]$, are given by

$$b = \inf_{\theta>0}[\frac{1}{\theta}(\ln\sum_{s=0}^{\infty}\mathrm{M}_A(\theta,s)\overline{\mathrm{M}}_S(\theta,s) - \ln\varepsilon)],$$

$$d = \inf_{\theta>0}\{\inf[\tau : \frac{1}{\theta}(\ln\sum_{s=\tau}^{\infty}\mathrm{M}_A(\theta,s-\tau)\overline{\mathrm{M}}_S(\theta,s) - \ln\varepsilon) \le 0]\}$$

where $\overline{\mathrm{M}}_S(\theta,0) = 1$ and for $s \ge 1$

$$\overline{\mathrm{M}}_S(\theta,s) = \breve{\boldsymbol{\pi}}(\mathbf{V}(-\theta)\breve{\mathbf{P}})^{s-1}\mathbf{V}(-\theta)\mathbf{1}_L, \tag{37}$$

where $\breve{\mathbf{P}}$ is determined by (34), $\breve{\boldsymbol{\pi}}$ by (30) and $\mathbf{V}(-\theta) = \mathrm{diag}(e^{-\theta r_1}, \cdots, e^{-\theta r_L})$.

TABLE I
SYSTEM PARAMETERS

Parameters	Values
Time unit (T_s)	2 ms
Number of states in FSMC (L)	5
Average SNR ($\bar{\gamma}$)	10 dB/12 dB
Maximum Doppler spread (f_D)	30 Hz/ 40 Hz
Source Period (N_p)	8
Arrival rate per user (λ_p)	(0,1)
Number of user (N)	$1 \sim 7$
Violation probability (ϵ)	10^{-2}

TABLE II
SNR THRESHOLDS AND RATES

State	Rate	MCS	$[\Gamma_l, \Gamma_{l+1})$
1	0	–	[0,1.7414)
2	1	QPSK, R_c=1/2	[1.7414,5.7187)
3	2	16QAM,R_c=1/2	[5.7187,13.8867)
4	3	16QAM,R_c=3/4	[13.8867, 31.908)
5	4	64QAM,R_c=2/3	[31.908,∞)

Note: R_c denotes the coding rate of convolutional Turbo code.

VI. NUMERICAL AND SIMULATION RESULTS

In this section, the delay and backlog performance of the multi-user FSMC is evaluated through both simulation and numerical results. Since *even sharing* is not implementable in real networks, the evaluation, if not otherwise specified, focuses on *exclusive use* channel sharing. In addition, since single-user is a special case of multi-user, its results are presented together.

For simplicity, a periodical source with the same parameters is assumed for each user. This source generates G packets at times $\{\alpha N_p + iN_p, i = 0, 1, 2, \cdots, \infty\}$ in every N_p time units of T_s, where the initial start time α is uniformly distributed over the interval of $[0, 1)$. The MGF of the corresponding arrival process is given by [19]

$$\mathrm{M}_A(\theta,t) = e^{\theta G\lfloor\frac{t}{N_p}\rfloor}[1 + (\frac{t}{N_p} - \lfloor\frac{t}{N_p}\rfloor)(e^{\theta G} - 1)]. \tag{38}$$

In our analysis and simulations, we set $N_p = 8$ and change the values of G, resulting in different packet arrive rates denoted by $\lambda_p = G/N_p$. The permissible violation probability of delay and backlog bound is set to be $\epsilon = 10^{-2}$.

The adopted parameters used in the evaluation, including the channel and system parameters, are summarized in Table I. Additionally for the FSMC channel, the signal-to-noise (SNR) thresholds and the corresponding service rates for a target block error rate (BLER) of 10^{-2} with the system parameters defined in IEEE 802.16m [44] are shown in Table II. Here, the service rate is represented by the number of packets that can be processed in a time unit of $T_s = 2ms$. The channel to each user is independent and identically distributed.

Our simulation program is built on the MATLAB platform. Each user has its buffer, where the arrival packets wait for being transmitted. At the start of each time unit, the scheduler allocates the radio resources to users according to the adopted channel sharing scheme, i.e., either *Method* 1 or *Method* 2. After the scheduling is carried out, the number of the

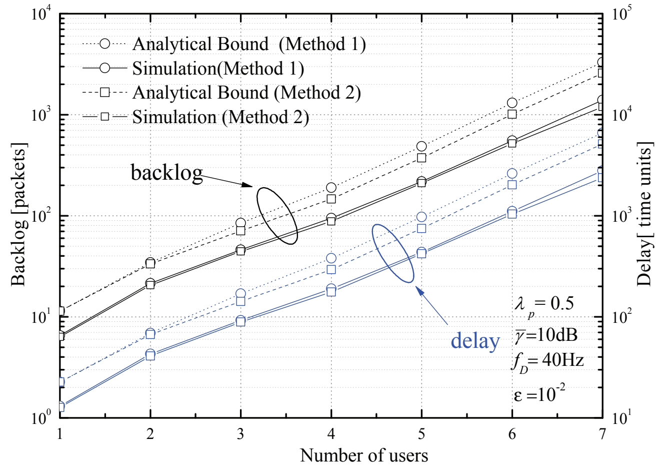


Fig. 1. Comparison of simulation results and analytical bounds for the two channel sharing methods ($\bar{\gamma}$=10dB, $f_D = 40$Hz).

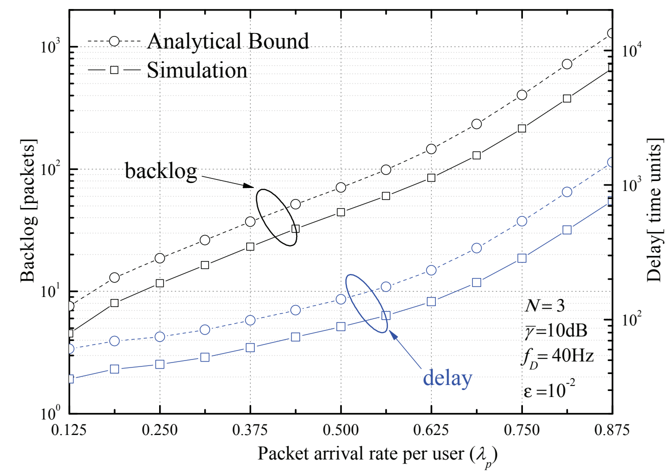


Fig. 2. Impact of packet arrival rate on delay and backlog performance ($\bar{\gamma}$=10dB, $f_D = 40$Hz, Method 2).

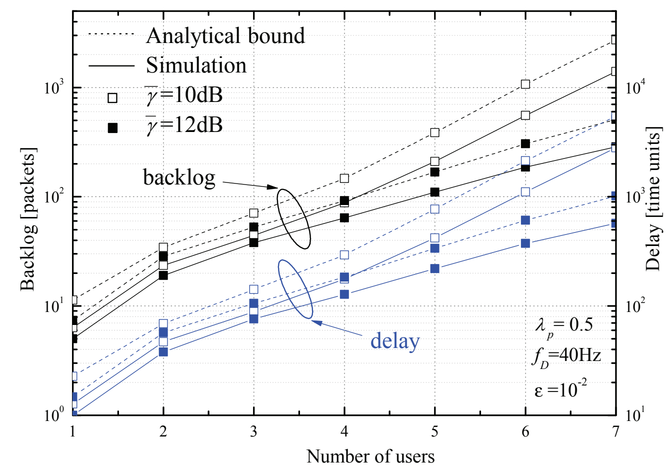


Fig. 3. Impact of SNR $\bar{\gamma}$ on delay and backlog performance ($f_D = 40$Hz, Method 2).

packets in each user's buffer is counted to analyze the backlog performance. Meanwhile, the sojourn time of each packet in buffer is recorded when the packet is transmitted. With this, the delay performance is obtained. The results are collected over 10^8 time periods for each parameter setting.

Fig. 1 compares the analytical bounds and the simulation results under both *even sharing (Method 1)* and *exclusive use (Method 2)* channel sharing, where the average received SNR $\bar{\gamma}$ and the Doppler spread are 10 dB and 40 Hz, respectively. Fig. 1 shows a good match between the analytical bounds and the corresponding simulation results and the former analytical bounds do bound the simulation results. This observation applies not only to the single-user case, but also to the multi-user case. In addition, for the single-user case, the two sharing methods are identical so are the analytical results as well as are the simulation results. This is also seen from the figure. Moreover, while Fig. 1 seems to show that one may use even-sharing analytical bounds to approximate for the practical exclusive-use sharing methods, the bounds from exclusive use analysis are tighter.

The alert reader may have noticed that we have highlighted in early discussion that *even sharing* is only a conceptual channel sharing method, not implementable in real networks. In obtaining simulation results for Fig. 1, we intentionally made it happen, i.e. gave each eligible user the equal portion[4] of the service rate and let all eligible users dequeue its traffic according to its share rate. The purpose is to give another reference for the comparison between analytical bounds and simulation results. It is hence worth highlighting that while Fig. 1 also includes simulation results for *even sharing*, this should not mislead the reader to think it is possible to implement *even sharing* in real networks. In the remaining, the comparison between analytical bounds and simulation results will be focused only on *exclusive use (Method 2)* channel sharing.

In Fig. 2, we compare analytical bounds and simulation results under different packet arrival rates, where the average SNR $\bar{\gamma}$ and the Doppler spread are set to 10 dB and 40 Hz, respectively, and the number of users is set to three. The figure shows that, with the increase of the packet arrival rate per user, the delay and backlog increase as expected: Since the network is essentially a queueing system where the service rate is limited by the wireless channel, the classic queueing theory tells that the delay and backlog increase with the system load. Fig. 2 also shows that the slope of the curves is steeper in the region of larger packet arrival rates, indicating that performance degradation is more severe in this region. This observation is also consistent with what may be predicted by the classic queueing theory, i.e., when the system is heavily loaded, a much faster increase in delay and backlog can generally be expected. In addition, Fig. 2 shows that the analytical results are in line with the simulation results, which demonstrates the effectiveness of the proposed analytical approach.

For the comparison in Fig. 3, we fix the packet arrival rate to be 0.5 and the Doppler spread $f_D = 40$ Hz, and investigate how the delay and backlog perform under different average

[4]This is one part making *even sharing* un-implementable in real networks, just like that the well-known generalized processor sharing (GPS) method is not implementable.

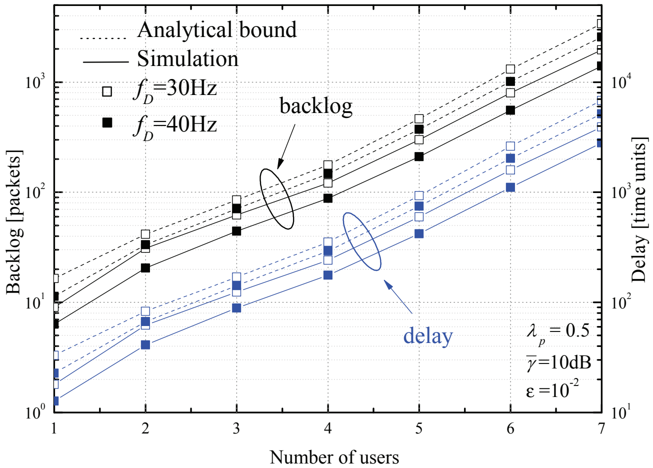


Fig. 4. Impact of Doppler spread on delay and backlog performance ($\bar{\gamma}$=10dB, Method 2).

SNRs $\bar{\gamma}$ and difference numbers of users. Specifically, results for $\bar{\gamma} = 10$ dB and $\bar{\gamma} = 12$ dB are plotted. Under the given parameter configurations, the stationary probability vector can be calculated as [0.1598, 0.2757, 0.3151, 0.2083, 0.0411] when $\bar{\gamma} = 10$ dB, while it is [0.1041, 0.1988, 0.2807, 0.2828, 0.1336] when $\bar{\gamma} = 12$ dB. It is clear from the figure that the higher the SNR, the better the delay and backlog performance. This is attributed to that, when the SNR $\bar{\gamma}$ increases, the channel is more likely (in total probability) to be in good states as demonstrated by the stationary probability vectors. This implies that the user is more likely to be served in time, reducing the delay and backlog in the system. Again, Fig. 3 demonstrates a close match between analytical bounds and simulation results.

Finally, Fig. 4 shows the impact of the Doppler spread on the delay and backlog performance under difference numbers of users. The packet arrival rate is set to be 0.5, and the SNR $\bar{\gamma}$ 10 dB. Results for Doppler spreads f_D=30 Hz and 40 Hz are plotted. Note that when the channel is in a poorer state, the service rate is smaller and thus the delay and backlog are increased. In contrast, if the channel is in a better state, a higher service rate is achievable, resulting in smaller delay and backlog. Therefore, the delay and backlog performance is contingent upon how long the channel is in deep fade that translates to poor channel states. With Doppler spread, an increase in f_D implies a shorter duration for the channel to stay in a poor state, implying better delay and backlog performance. This is also shown in Fig. 4, where the delay and backlog performance is better under f_D=40 Hz than under $f_D = 30$ Hz. The figure also shows that the analytical bounds well match with the simulation results.

VII. Discussion

In this paper, delay and backlog bounds have been derived and compared with simulation results for a wireless FSMC shared by multiple users. To help clarify the contributions of this work, we discuss in the following obtained results, the assumptions for obtaining those results, and potential extensions.

- From the formulation, Theorem 2 and Theorem 3 seem to resemble each other. However, it is worth emphasizing that there is a critical difference. The alert reader may have noticed that in Theorem 2, the state space size of $\tilde{\mathbf{P}}$, $\tilde{\boldsymbol{\pi}}$ and $\tilde{\mathbf{V}}_n(-\theta)$ are determined by L^N that is dependent on both the number of channel states in the FSMC model, i.e., L, and the number of users, i.e., N. When the number of users increases, the state space size in Theorem 2 increases exponentially, causing a tremendous computational burden in calculating the bounds. In contrast, in Theorem 3, with the proposed approach, the state space sizes of $\check{\mathbf{P}}$, $\check{\boldsymbol{\pi}}$ and $\mathbf{V}(-\theta)$ are determined only by L, and independent of N, resulting in a significantly reduced computational complexity.
- In the analysis, we have assumed that the channel experienced by each user is statistically independent from that of another. While this assumption seems to be somewhat restrictive, the fundamental idea of applying stochastic network calculus to performance analysis of the considered multi-user FSMC is demonstrated. In addition, we remark that FSMC is commonly used to model small-scale propagation, i.e., fast fading channel, and due to fast fading effect, we believe, the independence assumption is a reasonable starting point. For the more general situation where users may experience dependent channels, except for the large limit case under random channel access where each user can be decoupled from the others and hence each user approximately sees a separate dedicated channel independent of other users (e.g. [33] [45]), how to take the potential dependence into consideration with limited number of users sharing a channel, to the best of our knowledge, remains an open challenge and we leave it as our future work.
- Also in the analysis, we have assumed that users experience statistically identical channels. This assumption is mainly meant for ease of exposition. One may notice that in deriving the obtained bounds, the state transition probability matrix in (24) and the steady-state probability vector in (25) play a central role. It is worth highlighting that both (24) and (25) do not rely on the identical channel assumption. In other words, in cases when the channel seen by each user is inhomogeneous, given the corresponding (24) and (25), the analytical results obtained in this paper are readily extended to such cases. However, in those cases, finding the corresponding (24) and (25) may be challenging, which is out of the scope of this paper and should be studied case by case. For example, assume that there are three users, where two of them experiences Rayleigh fading channels and one of them goes through Rice fading channel independently, and, the same SNR threshold is applied to distinguish FSMC state sets, in which one is for Rayleigh fading channel and another is for Rice fading channel with the same number of states. In this case, the FSMC seen by users is inhomogeneous. Nevertheless, the state transition probabilities and consequently steady-state probabilities may still be able to find, with which, results in this paper can be extended, but for each user, they have to be derived individually since they may be different between users.

- Network calculus has a powerful result, known as concatenation property, which can be used for end-to-end performance guarantee analysis [16] [17] [18]. However, to apply this property, the single hop characterization of service process has to be studied first. For this reason, we have chosen to focus on single hop wireless FSMC channel. While this case is fundamental, it has surprisingly not been well studied. We believe, by making use of the stochastic network calculus concatenation property, the obtained results in this paper can be extended to multi-hop FSMC cases, but we leave this for future study.
- The performance of a multi-user FSMC depends on the scheduling strategy in use. In this paper, we have assumed the greedy algorithm [42]. We remark that while this greedy algorithm is optimal for capacity-approaching, it may not be so under other criteria. For example, the greedy algorithm does not ensure fairness among users. There are many other scheduling algorithms for wireless networks, which may perform better than the greedy algorithm under certain criteria [46]. It will be interesting to analyze their stochastic performance guarantees using stochastic network calculus, but to the best of our knowledge this remains largely open. A future work is hence to investigate how to extend to their analysis.

VIII. Conclusion

In this paper, delay and backlog bounds of a wireless FSMC shared by multiple users are derived. The derivation is based on the analytical principle behind the stochastic network calculus. When employing the MAX-SNR scheduling algorithm, both the *even sharing* and *exclusive use* methods for channel sharing are investigated. Particular focus is on the *exclusive use* method, which is more applicable in practical wireless networks. To address the state space exploration challenge in result computation, a novel method is adopted in the analysis for the *exclusive use* method. Its key idea is to used a modified state process, which has the same transition and stationary probabilities of states of interest, to approximate the original state process, and the modified process has unchanged state space size with the different user number. Numerical results and simulation results under various system parameter settings are presented and compared. The comparison indicates a good match between the analytical bounds and the simulation results, validating the analysis.

References

[1] H. Wang and N. Moayeri, "Finite-state Markov channel—a useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol. 44, no. 1, pp. 163–171, 1995.

[2] E. Gilbert and *et al.*, "Capacity of a burst-noise channel," *Bell Syst. Tech. J.*, vol. 39, no. 9, pp. 1253–1265, 1960.

[3] E. Elliott, "Estimates of error rates for codes on burst-noise channels," *Bell Syst. Tech. J.*, vol. 42, no. 9, pp. 1977–1997, 1963.

[4] Q. Zhang and S. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1688–1692, 1999.

[5] M. Zorzi and R. Rao, "On the statistics of block errors in bursty channels," *IEEE Trans. Commun.*, vol. 45, no. 6, pp. 660–667, 1997.

[6] M. Zorzi, R. Rao, and L. Milstein, "ARQ error control for fading mobile radio channels," *IEEE Trans. Veh. Technol.*, vol. 46, no. 2, pp. 445–455, 1997.

[7] H. Wang and P. Chang, "On verifying the first-order Markovian assumption for a Rayleigh fading channel model," *IEEE Trans. Veh. Technol.*, vol. 45, no. 2, pp. 353–357, 1996.

[8] Q. Liu, S. Zhou, and G. Giannakis, "Queuing with adaptive modulation and coding over wireless links: cross-layer analysis and design," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 1142–1153, 2005.

[9] L. Lei, C. Lin, J. Cai, and X. Shen, "Flow-level performance of opportunistic OFDM-TDMA and OFDMA networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5461–5472, 2008.

[10] M. Rashid, J. Hossain, E. Hossain, and V. Bhargava, "Opportunistic spectrum access in cognitive radio networks: a queueing analytic model and admission controller design," in *2007 IEEE GLOBECOM*.

[11] M. Luccini, A. Shami, and S. Primak, "Cross-layer optimisation of network performance over multiple-input multiple-output wireless mobile channels," *IET Commun.*, vol. 4, pp. 683–696, 2010.

[12] M. Fidler, "A network calculus approach to probabilistic quality of service analysis of fading channels," in *2006 IEEE GLOBECOM*.

[13] K. Mahmood, A. Rizk, and Y. Jiang, "On the flow-level delay of a spatial multiplexing MIMO wireless channel," in *2011 IEEE ICC*.

[14] R. Cruz, "A calculus for network delay–I: network elements in isolation," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 114–131, 1991.

[15] A. Burchard, J. Liebeherr, and S. Patek, "A min-plus calculus for end-to-end statistical service guarantees," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4105–4114, 2006.

[16] C. Chang, *Performance Guarantees in Communication Networks*. Springer, 2000.

[17] J. Le Boudec and P. Thiran, *Network Calculus: A Theory of Deterministic Queuing Systems for the Internet*. Springer-Verlag, 2001.

[18] Y. Jiang and Y. Liu, *Stochastic Network Calculus*. Springer, 2008.

[19] F. P. Kelly, "Notes on effective bandwidths," pp. 141–168, 1996.

[20] C. Li, A. Burchard, and J. Liebeherr, "A network calculus with effective bandwidth," *IEEE/ACM Trans. Networking*, vol. 15, no. 6, pp. 1442–1453, 2007.

[21] C. S. Chang, "Stability, queue length, and delay of deterministic and stochastic queueing networks," *IEEE Trans. Auto. Control*, pp. 913–931, 1994.

[22] Q. Yin, Y. Jiang, S. Jiang, and P. Y. Kong, "Analysis on generalized stochastically bounded bursty traffic for communication networks," in *2002 IEEE LCN*.

[23] M. Fidler, "An end-to-end probabilistic network calculus with moment generating functions," in *2006 IEEE International Workshop on Quality of Service*.

[24] Y. Jiang, "A basic stochastic network calculus," in *Proc. 2006 ACM SIGCOMM*, pp. 123–134.

[25] Y. Liu, C. Tham, and Y. Jiang, "A calculus for stochastic QoS analysis," *Performance Evaluation*, vol. 64, pp. 547–572, 2007.

[26] Y. Jiang, "Network calculus and queueing theory: two sides of one coin," in *2009 International Conference on Performance Evaluation Methodologies and Tools*.

[27] Y. Jiang, Q. Yin, Y. Liu, and S. Jiang, "Fundamental calculus on generalized stochastically bounded bursty traffic for communication networks," *Computer Networks*, vol. 53, no. 12, pp. 2011–2021, 2009.

[28] Y. Yuan, K. Wu, W. Jia, and Y. Jiang, "Performance of acyclic stochastic networks with network coding," *IEEE Trans. Parallel Distrib. Syst.*, vol. 22, no. 7, pp. 1238–1245, 2011.

[29] K. Zheng, L. Lei, Y. Wang, Y. Lin, and W. Wang, "Quality-of-service performance bounds in wireless multi-hop relaying networks," *IET Commun.*, vol. 5, pp. 71–78, 2011.

[30] H. A. Zubaidy, J. Liebeherr, and A. Burchard, "A network calculus approach for the analysis of multi-hop fading channels," arXiv:1207.6630v1 [cs.NI], July 2012.

[31] D. Wu and R. Negi, "Effective capacity: a wireless link model for support of quality of service," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 630–643, 2003.

[32] A. Stolyar, "Large deviations of queues sharing a randomly time-varying server," *Queueing Systems*, vol. 59, no. 1, pp. 1–35, 2008.

[33] A. Kumar, E. Altman, D. Miorandi, and M. Goyal, "New insights from a fixed point analysis of single cell ieee 802.11 WLANs," in *Proc. 2005 IEEE INFOCOM*, vol. 3, pp. 1550–1561.

[34] S. Bodas, S. Shakkottai, L. Ying, and R. Srikant, "Scheduling for small delay in multi-rate multi-channel wireless networks," in *Proc. 2011 IEEE INFOCOM*, pp. 1251–1259.

[35] F. Agharebparast and V. Leung, "Modeling wireless link layer by network for efficient evaluations of multimedia QoS," in *2005 IEEE ICC*.

[36] Y. Jiang and P. Emstad, "Analysis of stochastic service guarantees in communication networks: a server model," in *2005 International Workshop on Quality of Service*.

[37] M. Fidler, "Survey of deterministic and stochastic service curve models in the network calculus," *IEEE Commun. Surveys & Tutorials*, vol. 12, no. 1, pp. 59–86, 2010.

[38] K. Mahmood, M. Vehkapera, and Y. Jiang, "Delay constrained throughput analysis of a correlated MIMO wireless channel," in *2011 IEEE ICCCN*.

[39] Y. Gao and Y. Jiang, "Analysis on the capacity of a cognitive radio network under delay constraints," *IEICE Trans. Commun.*, vol. E95-B, no. 4, pp. 1180–1189, 2012.

[40] F. Ciucu, "On the scaling of non-asymptotic capacity in multi-access networks with bursty traffic," in *Proc. 2011 IEEE International Symposium on Information Theory*, pp. 2547–2551.

[41] S. Wang, J. Zhang, and L. Tong, "Delay analysis for cognitive radio networks with random access: a fluid queue view," in *2010 IEEE INFOCOM*.

[42] R. Knopp and P. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. 1995 IEEE International Conference on Communications*, vol. 1, pp. 331–335.

[43] T. Stern and A. Elwalid, "Analysis of separable Markov-modulated rate models for information-handling systems," *Advances in Applied Probability*, pp. 105–139, 1991.

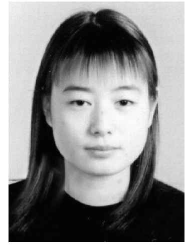
[44] R. Srinivasan, J. Zhuang, L. Jalloul, R. Novak, and J. Park, "IEEE 802.16 m evaluation methodology document (EMD)," IEEE 802.16 Broadband Wireless Access Working Group, 2008.

[45] K. Mahmood, M. Vehkapera, and Y. Jiang, "Performance of multiuser CDMA receivers with bursty traffic and delay constraints," in *Proc. 2012 IEEE International Conference on Computing, Networking and Communications*.

[46] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

Kan Zheng (SM'09) received the B.S., M.S. and Ph.D. degree from Beijing University of Posts&Telecommunications (BUPT), China, in 1996, 2000 and 2005, respectively, where he is currently associate professor. He worked as a researcher in the companies including Siemens, Orange Labs R&D (Beijing), China. His current research interests lie in the field of wireless communications, with emphasis on heterogeneous networks and M2M networks.

Fei Liu received his B.S. degree from the School of Information and Communication Engineering, Beijing University&Posts and Telecommunications (BUPT), China, in 2010. He is currently a candidate for M.S. in the Key Lab of Universal Wireless Communications, Ministry of Education, BUPT. His research interests include performance analysis of wireless networks, resource allocation and scheduling algorithm.



Lei Lei received a B.S. degree in 2001 and a Ph.D. degree in 2006, respectively, from Beijing University of Posts & Telecommunications, China, both in telecommunications engineering. She is currently associate professor in Beijing Jiaotong University. Her current research interests include performance evaluation, quality-of-service and radio resource management in wireless communication networks.

Chuang Lin (SM'04) is a professor of the Department of Computer Science and Technology, Tsinghua University, Beijing, China. He received the Ph.D. degree in Computer Science from the Tsinghua University in 1994. His current research interests include computer networks, performance evaluation, network security analysis, and Petri net theory and its applications. He has published more than 300 papers in research journals and IEEE conference proceedings in these areas and has published three books.

Professor Lin is a member of ACM Council, a senior member of the IEEE and the Chinese Delegate in TC6 of IFIP. He serves as the Technical Program Vice Chair, the 10th IEEE Workshop on Future Trends of Distributed Computing Systems (FTDCS 2004); the General Chair, ACM SIGCOMM Asia workshop 2005; the Associate Editor, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY; the Area Editor, *Journal of Computer Networks*; and the Area Editor, *Journal of Parallel and Distributed Computing*.

Yuming Jiang received his BSc from Peking University, China, in 1988, MEng from Beijing Institute of Technology, China, in 1991, and Ph.D. from National University of Singapore, Singapore, in 2001. He worked with Motorola from 1996 to 1997. From 2001 to 2003, he was a Member of Technical Staff and Research Scientist with the Institute for Infocomm Research, Singapore. From 2003 to 2004, he was an Adjunct Assistant Professor with the Electrical and Computer Engineering Department, National University of Singapore. From 2004 to 2005, he was with the Centre for Quantifiable Quality of Service in Communication Systems (Q2S), Norwegian University of Science and Technology (NTNU), Norway, supported in part by the Fellowship Programme of European Research Consortium for Informatics and Mathematics (ERCIM). Since 2005, he has been with the Department of Telematics, NTNU, as a Professor. He visited Northwestern University, USA from 2009 to 2010.

He was Co-Chair of IEEE Globecom2005 - General Conference Symposium, TPC Co-Chair of 67th IEEE Vehicular Technology Conference (VTC) 2008, and General/TPC Co-Chair of International Symposium on Wireless Communication Systems (ISWCS) 2007-2010. He is first author of the book *Stochastic Network Calculus*. His research interests are the provision, analysis and management of quality of service guarantees in communication networks. In the area of network calculus, his focus has been on developing models and investigating their basic properties for stochastic network calculus (snetcal), and recently also on applying snetcal to performance analysis of wireless networks.